TWNFC – Transductive Neural-Fuzzy Classifier with Weighted Data Normalization and Its Application in Medicine

T. M. Ma[1, 2], Q. Song[1], M. R. Marshall[2,3], N. Kasabov[1]

[1]*Knowledge Engineering & Discovery Research Institute*
Auckland University of Technology
Private Bag 92006, Auckland 1020, New Zealand
E-mail: mmaa@aut.ac.nz, qsong@aut.ac.nz, nkasabov@aut.ac.nz
[2]*Department of Renal Medicine, Middlemore Hospital*
Private Bag 93311, Auckland, New Zealand
[3]*The Dialysis Outcomes and Practice Patterns Study (DOPPS)*
E-mail: mrmarshall@middlemore.co.nz

Abstract

This paper introduces a novel fuzzy model – transductive neural-fuzzy classifier with weighted data normalization (TWNFC). While inductive approaches are concerned with the development of a model to approximate data in the whole problem space (induction), and consecutively – using this model to calculate the output value(s) for a new input vector (deduction), in transductive systems a local model is developed for every new input vector, based on some closest data to this vector from the training data set. The weighted data normalization method (WDN) optimizes the data normalization ranges for the input variables of a system. A steepest descent algorithm is used for training the TWNFC model. The TWNFC is illustrated on a case study: a real medical decision support problem of estimating the survival of haemodialysis patients. This personalized modeling can also be applied to other distance-based, prototype learning neural network or fuzzy inference models.

1. Introduction: transductive model and weighted data normalization

Most of learning models and systems in artificial intelligence developed and implemented so far are based on inductive methods, where a model (a function) is derived from data representing the problem space and subsequently applied on new data. The derivation of the model in this manner therefore may not optimally account for all of the specific information related to a given new vector in the test data. An error is measured to estimate how well the new data fits into the model. The inductive learning and inference approach is useful when a global model ("the big picture") of the problem is needed. In contrast, transductive inference methods estimate the value of a potential model (function) only in a single point of the space (the new data vector) utilizing additional information related to this point. This approach seems to be more appropriate for medical applications, where the focus is not on the model, but on the individual patient. Each individual data vector (e.g.: a patient in the medical area; a future time moment for predicting time series; or a target day for predicting a stock index) may need an individual, local model that fits the new data better than a global model, in which the new data is matched without taking any specific information about this data into account [1,2].

Transductive inference is concerned with the estimation of a function in single point of the space only. For every new input vector xi that needs to be processed for a prognostic task, the Ni nearest neighbours, which form a sub-dataset Di, are derived from an existing data set D and, if necessary, generated from an existing model M. A new model Mi is dynamically created from these samples to approximate the function in the point xi - Figure 1 and Figure 2. Then the system is used to calculate the output value yi for this input vector xi (Figure 1 and 2).

In many neural networks, fuzzy models and their applications, raw data without normalization is used. This is appropriate when all the input variables are measured in the same units. Normalization, or standardization, is reasonable when the variables are in different units, or when the variance between them is

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe
Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)
0-7695-2504-0/05 $20.00 © 2005 IEEE

substantial. However, a general normalization means that every variable is normalized in the same range, e.g. [0, 1] with the assumption that they all have the same importance for the output of the system.

For many practical problems, variables have different importance and make different contribution to the output(s). Therefore, it is necessary to find an optimal normalization and assign proper importance factors to the variables. Such a method can also be used for feature selection or for reducing the size of input vectors through keeping the most important ones [3]. This is especially applicable to a special class of neural networks or fuzzy models – the clustering based (or distance-based; prototype-based) models such as radial based function (RBF) [4] and Evolving Connectionist System (ECOS) [5,6]. In such systems, distance between neurons or fuzzy rule nodes and input vectors are usually measured in Euclidean distance, so that variables with wider ranges will have more influence on the learning process and vice versa.

The paper is organized as follows: Section II presents the structure and algorithm of the TWNFC model. Section III illustrates the approach on a case study example. Conclusions are drawn in Section IV.

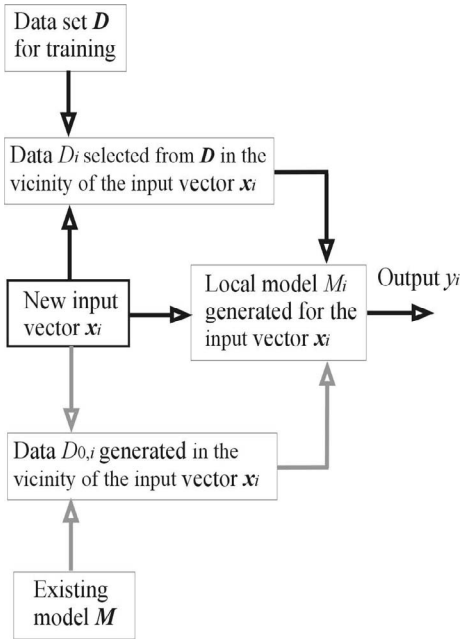
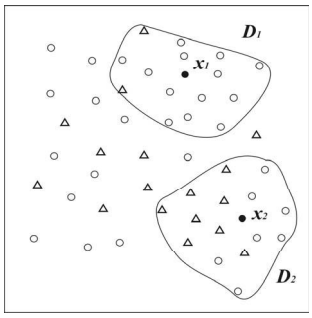


Figure 1. A block diagram of a transductive reasoning system. An individual model $\boldsymbol{M}_i$ is trained for every new input vector $\boldsymbol{x}_i$ with the use of data samples $\boldsymbol{D}_i$ selected from a data set $\boldsymbol{D}$, and data samples $\boldsymbol{D}_{\theta i}$ generated from an existing model (formula) $\boldsymbol{M}$ (if such a model is existing). Data samples in both $\boldsymbol{D}_i$ and $\boldsymbol{D}_{0,i}$ are similar to the new vector $\boldsymbol{x}_i$ according to defined similarity criteria.



● – a new data vector; ○ – a sample from $\boldsymbol{D}$; Δ – a sample from $\boldsymbol{M}$

Figure 2. In the centre of a transductive reasoning system is the new data vector (here illustrated two of them – $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$), surrounded by a fixed number of nearest data samples selected from the training data set $\boldsymbol{D}$ and generated from an existing model $\boldsymbol{M}$.

2. Transductive Neural Fuzzy Systems with weighted data normalization: structure and learning algorithm

TWNFC is a dynamic neural-fuzzy inference system with a local generalization, in which, the Zadeh-Mamdani type fuzzy inference engine is used [7]. The local generalization means that in a sub-space (local area) of the whole problem space, a model is created, which performs generalization in this area. In the TWNFC model, Gaussian fuzzy membership functions are applied in each fuzzy rule for both antecedent and consequent parts. A steepest descent back- propagation (BP) learning algorithm is used for optimizing the parameters of the fuzzy membership functions [8,9]. The distance between vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is measured in TWNFC in normalized Euclidean distance defined as follows (the values are between 0 and 1):

$$\|\boldsymbol{x}-\boldsymbol{y}\| = \frac{1}{P}\left[\sum_{j=1}^{P}|x_j-y_j|^2\right]^{\frac{1}{2}} \quad (1)$$

(1)

where: $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{R}^P$

To partition the input space for creating fuzzy rules and obtaining initial values of fuzzy rules, the Evolving Clustering Method (ECM) is applied [10,11] and the cluster centers and cluster radiuses are respectively taken as initial values of the centers and widths of the Gaussian membership functions.

For each new data vector $\boldsymbol{x}_q$, the TWNFC learning algorithm performs the following steps:

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe
Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)
0-7695-2504-0/05 $20.00 © 2005 IEEE

1) Normalize the training data set (the values are between 0 and 1) with the initial weights of input variables.
2) Search in the training data set in the input space to find N_q training examples that are closest to x_q. The value for N_q can be pre-defined based on experience, or - optimized through the application of an optimization procedure. Here we assume the former approach.
3) Calculate the distances d_i, i = 1, 2, …, N_q, between x_q and each of these N_q data samples . Calculate the vector weights v_i = 1 – (d_i – min(d)), i = 1, 2, …, N_q, min(d) is the minimum value in the distance vector d, d = [d_1, d_2, … , d_{Nq}].
4) Use the ECM clustering algorithm to cluster and partition the input sub-space that consists of N_q selected training samples.
5) Create fuzzy rules and set their initial parameter values according to the results of ECM clustering procedure. For each cluster, the cluster centre is taken as the centre of a fuzzy membership function (Gaussian function) and the cluster radius is taken as the width.
6) Apply the steepest descent method (Bp) to optimize the parameters of the fuzzy rules in the local model M_q following Eq. (6 – 13).
7) Re-normalize the training data set (the values are between 0 and 1) with the optimized weights of variables.
8) Search in the training data set to find N_q nearest samples (same to Step 2), if the same samples are found, as the last search, the algorithm turns to Step 9, otherwise, Step 3.
9) Calculate the output value y_q for the input vector x_q applying fuzzy inference over the set of fuzzy rules that constitute the local model M_q.
10) End of the procedure.

The parameter optimization procedure is described below:

Consider the system having P inputs, one output and M fuzzy rules defined initially through the ECM clustering procedure, the l-th rule has the form of:

R_l : If x_1 is F_{l1} and x_2 is F_{l2} and … x_P is F_{lP}, then y is G_l. (2)

Here, F_{lj} are fuzzy sets defined by the following Gaussian type membership function:

$$GaussianMF = \alpha \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] \quad (3)$$

G_l are of a similar type as F_{lj} and are defined as:

$$GaussianMF = \exp\left[-\frac{(y-n)^2}{2\delta^2}\right] \quad (4)$$

Using the modified centre average defuzzification procedure, the output value of the system can be calculated for an input vector $x_i = [x_1, x_2, \ldots, x_P]$ as follows:

$$f(x_i) = \frac{\sum_{l=1}^{M} \frac{G_l}{\delta_l^2} \prod_{j=1}^{P} \alpha_{lj} \exp\left[-\frac{w_j^2 (x_{ij} - m_{lj})^2}{2\sigma_{lj}^2}\right]}{\sum_{l=1}^{M} \frac{1}{\delta_l^2} \prod_{j=1}^{P} \alpha_{lj} \exp\left[-\frac{w_j^2 (x_{ij} - m_{lj})^2}{2\sigma_{lj}^2}\right]} \quad (5)$$

Here, w_j are weights of the input variables.

Suppose a training input-output data pair [x_i, t_i] is given to the TWNFC, the system minimizes the following objective function (a weighted error function):

$$E = \frac{1}{2} v_i \left[f(x_i) - t_i\right]^2 \quad (6)$$

(v_i are defined in Step 3)

Then the steepest descent algorithm (BP) is used to obtain the formulas for the optimization of the parameters G_l, δ_l, α_{lj}, m_{lj}, σ_{lj} and w_j, so that the value of E from Eq. (6) is minimized:

$$G_l(k+1) = G_l(k) - \frac{\eta_G}{\delta_l^2(k)} v_i \Phi(x_i) \left[f^{(k)}(x_i) - t_i\right] \quad (7)$$

$$\delta_l(k+1) = \delta_l(k) - \frac{\eta_\delta v_i \Phi(x_i)}{\delta_l^3(k)} \left[f^{(k)}(x_i) - t_i\right]\left[f^{(k)}(x_i) - G_l(k)\right] \quad (8)$$

$$\alpha_{lj}(k+1) = \alpha_{lj}(k) - \frac{\eta_\alpha v_i \Phi(x_i)}{\delta_l^2(k)\alpha_{lj}(k)} \left[f^{(k)}(x_i) - t_i\right]\left[G_l(k) - f^{(k)}(x_i)\right] \quad (9)$$

$$m_{lj}(k+1) = m_{lj}(k) - \frac{\eta_m w_j^2(k) v_i \Phi(x_i)}{\delta_l^2(k)\sigma_{lj}^2(k)} \left[f^{(k)}(x_i) - t_i\right]\left[G_l(k) - f^{(k)}(x_i)\right]\left[x_{ij} - m_{lj}(k)\right] \quad (10)$$

$$c_{lj}(k+1) = c_{lj}(k) - \frac{\eta_\sigma w_j^2(k) v_i \Phi(x_i)}{\delta_l^2(k)\,\sigma_{lj}^3(k)} \left[f^{(k)}(x_i) - t_i\right]\left[G_l(k) - f^{(k)}(x_i)\right]\left[x_{ij} - m_{lj}(k)\right]^2 \quad (11)$$

$$w_j(k+1) = w_j(k) - \frac{\eta_w w_j(k) v_i \Phi(x_i)}{\delta_l^2(k)\sigma_{lj}^2(k)} \left[f^{(k)}(x_i) - t_i\right]\left[f^{(k)}(x_i) - G_l(k)\right]\left[x_{ij} - m_{lj}(k)\right]^2 \quad (12)$$

$$\Phi(x_i) = \frac{\prod_{j=1}^{P} \alpha_{lj} \exp\left\{-\frac{w_j^2(k)\left[x_{ij} - m_{lj}(k)\right]^2}{2\sigma_{lj}^2(k)}\right\}}{\sum_{l=1}^{M} \frac{1}{\delta_l^2} \prod_{j=1}^{P} \alpha_{lj} \exp\left\{-\frac{w_j^2(k)\left[x_{ij} - m_{lj}(k)\right]^2}{2\sigma_{lj}^2(k)}\right\}}$$

(13)

η_G, η_δ, η_α, η_m, η_σ and η_w are learning rates for updating the parameters G_l, δ_l, α_{lj}, m_{lj}, σ_{lj} and w_j respectively.

In the TWNFC training–simulating algorithm, the following indexes are used:

- Training data samples: $i = 1, 2, \ldots, N$;
- Input variables: $j = 1, 2, \ldots, P$;
- Fuzzy rules: $l = 1, 2, \ldots, M$;
- Training epochs: $k = 1, 2, \ldots$.

3. Case study example of applying the TWNFC for a medical decision support problem

A medical dataset is used here for experimental analysis. Data originate from the Dialysis Outcomes and Practice Patterns Study (DOPPS, www.dopps.org) [12]. The DOPPS is based upon the prospective collection of observational longitudinal data from a stratified random sample of haemodialysis patients from the United Sates, 8 European countries (United Kingdom, France, Germany, Italy, Spain, Belgium, Netherlands, and Sweden), Japan, Australia and New Zealand. There have been two phases of data collection since 1996, and a third phase is currently just beginning. To date, 27,880 incident and prevalent patients (approximately 33% and 66% respectively) have been enrolled in the study, which represents approximately 75% of the world's haemodialysis patients. In this study, prevalent patients are defined as those patients who had received maintenance hemodialysis prior to the study period, while incident patients are those who had not previously received maintenance hemodialysis.

The research plan of the DOPPS is to assess the relationship between haemodialysis treatment practices and patient outcomes. Detailed practice pattern data, demographics, cause of end-stage renal disease, medical and psychosocial history, and laboratory data are collected at enrollment and at regular intervals during the study period. Patient outcomes studied include mortality, frequency of hospitalisation, vascular access, and quality of life. The DOPPS aims to measure how a given practice changes patient outcomes, and also determine whether there is any relationship amongst these outcomes, for the eventual purpose of improving treatments and survival of patients on haemodialysis.

The dataset for this case study contains 6100 samples from the DOPPS phase 1 in the United States, collected from 1996-1999. Each record includes 24 patient and treatment related variables (input): demographics (age, sex, race), psychosocial characteristics (mobility, summary physical and mental component scores (sMCS, sPCS) using the Kidney Disease Quality of Life (KD-QOL®) Instrument), co-morbid medical conditions (diabetes, angina, myocardial infarction, congestive heart failure, left ventricular hypertrophy, peripheral vascular disease, cerebrovascular disease, hypertension, body mass index), laboratory results (serum creatinine, calcium, phosphate, albumin, hemoglobin), haemodialysis treatment parameters (Kt/V, haemodialysis angioaccess type, haemodialyser flux), and vintage (years on haemodialysis at the commencement of the DOPPS). The output is survival at 3 years from study enrollment (yes or no). All experimental results reported here are based on 10-cross validation experiments [13].

For comparison, several well-known methods of classification are applied to the same problem, such as Support Vector Machine (SVM) [14], Evolving Classification Function (ECF) [5,15], Multi-Layer Perceptron (MLP) [15], Radial Basis Function (RBF) [15], and Multiple Linear Regression along with the proposed TWNFC, and results are given in Table 1.

The Kappa statistic, K, formally tests for agreement between two methods, raters, or observers, when the observations are measured on a categorical scale. Both methods must rate, or classify, the same cases using the same categorical scale [16]. The degree of agreement is indicated by K, which can be roughly interpreted as follows: $K < 0.20$, agreement quality poor; $0.20 < K < 0.40$, agreement quality fair; $0.40 < K < 0.60$, agreement quality moderate; $0.60 < K < 0.80$, agreement quality good; $K > 0.80$, agreement quality very good. Confidence intervals for K were constructed using the goodness-of-fit approach of Donner & Eliasziw [17]. There is no universally agreed method for comparing K between multiple tests of agreement. In this study, K for different classification methods was compared using the permutation or Monte Carlo resampling routine of McKenzie [18,19].

Agreement refers to the quality of the information provided by the classification device and should be distinguished from the usefulness, or actual practical value, of the information. Agreement provides a pure index of accuracy by demonstrating the limits of a test's ability to discriminate between alternative states of health over the complete spectrum of operating conditions. To date, prognostic systems for the prediction of haemodialysis patient survival have published accuracy of 60-70%. The experimental results in Table 1 illustrate that the TWNFC in this paper provide incrementally better results, towards a K of > 0.60 and a level of accuracy ~80% ,which are generally regarded as thresholds for clinical utility.

For every patient sample, a personalised model will

be created and used to evaluate the output value for the patient, and also to estimate the importance of the variables for this patient using Equation (12). Two examples are shown in Table 2. The TWNFC not only results in a better accuracy for these patients, but also shows the importance of the variables for her/him that may result in a more efficient personalised treatment.

Table 1. Experimental results on the DOPPS data

Model	Kappa (95% Confidence Intervals)*	P-value	Agreement (%)	Specificity (%)	Sensitivity (%)
RBF	0.1675 (0.1268 - 0.2026)	<0.001	59.1	67.51	49.08
ECF	0.1862 (0.1469 - 0.2224)	<0.001	59.9	66.74	51.76
MLP	0.3833 (0.3472 - 0.4182)	<0.001	69.44	72.56	65.72
Multiple Linear Regression	0.4020 (0.3651 - 0.4357)	<0.001	70.55	76.7	63.21
SVM	0.4110 (0.3748 - 0.4449)	<0.001	70.93	76	64.88
TWNFC	0.4503 (0.4152 - 0.4837)	Reference	72.64	73.3	71.8

- Kappa values and confidence intervals ascertained with Stata Intercooled V 8.2 (StataCorp, College Station, TX), and P-values with KAPCOM [19]

Table 2. TWNFC models of single patient (two samples from the DOPPS data)

	Patient 1		Patient 2	
Input variables	Values of input	Weights of input variables	Values of input	Weights of input variables
Years on Dialysis prior to Study	0.34	0.49	0.5175	0.63
Age	88	0.85	66	1
Sex	Female	0.05	Female	0.62
Race	Black	0.59	White	0.72
Diabetes	No	0.96	No	0.56
Angina	Angina at rest within 12 months of enrolment date	1	No	0.89
Myocardial Infarction	Yes	0.77	No	0.62
Chronic Heart Failure	Dyspnea at rest or pulmonary edema	0.54	No	0.71
Left Ventricular Hypertrophy	Yes	0.79	No	0.33
Serum Albumin	3.8667	0.54	3.7	0.94
Peripheral Vascular Disease	No	0.37	No	0.68
Cerebrovascular Disease	No	0.73	No	0.21
Hypertension	Yes	0.76	Yes	0.7
Kt/V	1.3	0.52	1.31	0.68
Serum Phosphate	4.9333	0.56	3.77	0.57
Serum Hemoglobin	11.3333	0.42	9.9	0.66
Type of access for Dialysis	Synthetic graft	0.95	Native A- V fistula	0.24
Mobility	Can walk with assistance	0.69	Can walk without assistance	0.5
sPCS	32.02	0.98	51.82	0.64
sMCS	50.99	0.77	43.99	0.69
Body Mass Index	23.5	0.6	17.5	0.6
Hi-flux	No	0.82	Yes	0.66
Serum Creatinine	6.8	0.6	5.93	0.8
Serum Calcium	8.53	0.52	9.07	0.6
Output	Survive	Predicting result: Survive	Non-survive	Predicting result: Non-survive

4. Conclusions

This paper presents a transductive neuro-fuzzy classifier with weighted data normalization method – TWNFC. The TWNFC performs a better local generalisation over new data as it develops an individual model for each data vector that takes the location of new input vector in the space into account. This approach seems to be more appropriate for clinical and medical applications of learning systems, where the focus is not on the model, but on the individual patient. At the same time, it is an adaptive model, in the sense that input-output pairs of data can be added to the data set continuously and immediately, and made available for transductive inference of local models. This type of modelling can be called "personalised", and it is promising for medical decision support systems. The clinical plausibility of the approach and its results are satisfactory in this study. As the TWNFC creates a unique sub-model for each data sample, it usually needs more performing time than inductive models, especially when training and simulating are based on large data sets.

Further directions for research include: (1) TWNFC system parameter optimization such as optimal number of nearest neighbours; and (2) applying the TWNFC method to other decision support systems, such as: cardio-vascular risk prognosis; biological processes modelling and classifications based on gene expression micro-array data.

5. Acknowledgement

The research presented in the paper is funded by the New Zealand Foundation for Research, Science and Technology under grant NERF/AUTX02-01, the New Zealand National Kidney Foundation under grant GSN-12, and the Auckland Medical Research Foundation. The Top Achiever Doctoral Scholarship from the Tertiary Education Commission (TEC) of New Zealand also funds the research.

6. References

[1] A. Gammerman, V. Vovk and V. Vapnik, "Learning by transduction," *Proc. of the 14th Conference on Uncertainty in Artificial Intelligence,* G. F. Cooper & S. Moral (Ed.), Madison, Wisconsin, Morgan Kaufmann, San Francisco, USA, 1998, pp. 148-155.

[2] M. Kukar, "Transductive reliability estimation for medical diagnosis," *Artif. Intell. Med.,* 2003, Vol. 29, pp. 81-106.

[3] Q. Song and N. Kasabov, "Weighted data normalizations and feature selection for Evolving Connectionist Systems Proceedings," *Proc. of the Eighth Australian and New Zealand Intelligence Information Systems Conference (ANZIIS2003),* B.C. Lovell, et al. (Ed.), University of Queensland, Sydney, Australia, 2003, pp. 285 – 290.

[4] F. Poggio, "Regularization theory, radial basis functions and networks," In: From Statistics to Neural Networks: Theory and Pattern Recognition Applications. *NATO ASI Series,* 1994, Vol. 136, pp. 83 – 104.

[5] N. Kasabov, *Evolving connectionist systems: Methods and Applications in Bioinformatics, Brain study and intelligent machines,* Springer Verlag London Limited, Great Britain, 2003.

[6] N. Kasabov, "Evolving fuzzy neural networks for on-line supervised/unsupervised, knowledge–based learning," *IEEE Trans. SMC – part B, Cybernetics,* 2001, Vol. 31, No. 6 , pp. 902 – 918.

[7] L. A. Zadeh, "Fuzzy Sets," *Information and Control,* 1965, Vol. 8, pp. 338 – 353.

[8] C.T. Lin and C.S.G.Lee, *Neuro Fuzzy Systems*, Prentice Hall, 1996.

[9] L.X. Wang, *Adaptive Fuzzy System And Control: Design and Stability Analysis*, Prentice Hall, Englewood Cliffs, NJ, 1994.

[10] N. Kasabov and Q. Song, "DENFIS: Dynamic, evolving neural-fuzzy inference systems and its application for time-series prediction," *IEEE Trans. on Fuzzy Systems,* 2002, Vol. 10, pp. 144 – 154.

[11] Q. Song and N. Kasabov, "ECM - A Novel On-line, Evolving Clustering Method and Its Applications," *Proc. of the Fifth Biannual Conference on Artificial Neural Networks and Expert Systems (ANNES2001),* University of Otago, Dunedin, New Zealand, 2001, pp. 87 – 92.

[12] D.A. Goodkin, D.L. Mapes and P.J. Held, "The dialysis outcomes and practice patterns study (DOPPS): how can we improve the care of hemodialysis patients?" *Seminars in Dialysis*, 2001, Vol. 14, pp. 157-159.

[13] M. R. Marshall, Q. Song, T.M. Ma, S. MacDonell and N. Kasabov, "Evolving Connectionist System versus Algebraic Formulae for Prediction of Renal Function from Serum Creatinine," *Kidney International,* 2005, Vol. 67, pp. 1944-1954.

[14] V. Vapnik, *The Nature of Statistical Learning Theory,* Springer-Verlag, New York, NY, 1995.

[15] *Neural Network Toolbox User's Guide*, The Math Works Inc., 3 Apple Hill Drive, Natrick, Massachusetts, Ver. 4 , 2002.

[16] D. G. Altman, *Practical Statistics for Medical Research,* Chapman and Hall, London, Great Britian, 1991

[17] A. Donner and M. Eliasziw, "A goodness-of-fit approach to inference procedures for the kappa statistic: confidence interval construction, significance-testing and sample size estimation," *Statistics in Medicine*, 1992, Vol 11:, pp. 1511-1519.

[18] D. P. McKenzie, A. J.Mackinnon, N. Peladeau, P. Onghena, P. C. Bruce, D. M. Clarke, S. Harrigan, and P. D. McGorry, "Comparing correlated kappas by resampling: is one level of agreement significantly different from another?" *Journal of Psychiatric Research*, 1996, Vol 30, pp. 483-492.

[19] D. P. McKenzie, A. J. Mackinnon, and D. M. Clarke, "KAPCOM: a program for the comparison of kappa coefficients obtained from the same sample of observations," *Perceptual and Motor Skills*, 1997, Vol 85, pp. 899-902.